Supplement of

Advancing climate model evaluation and intercomparison: an error decomposition-based index to quantify statistical sources of climate model errors

Meng-Zhuo Zhang et al.

Correspondence to: Zhongfeng Xu (xuzhf@tea.ac.cn)

Text

Text S1

Calculation formulas for performance metrics in the multivariable integrated evaluation (MVIE) method are provided first. Assume that there are *M* individual variables comprising either scalar (e.g., temperature, specific humidity) or vector fields (e.g., vector wind) from the model and observation (Xu et al., 2017; Zhang et al., 2021). Therein, the field of *m*-th variable features the dimension of $d_m \times N$, where N denotes the num of time or/and space grids. d_m is 1 for a scalar variable, but this value is greater than 1 for a vector variable, such as 2 for 2D wind fields. Normalize each scalar (vector) variable for both model and observation by the root mean square value (root mean square length) of the corresponding observed one, and then group *M* individual variables into multiple fields featuring the dimension $D \times N$, where D is the sum of d_m across all evaluated variables. We express the normalized multiple fields derived from the model and observation as:

$$\boldsymbol{A_j} = \left(a_{1j}, a_{2j}, \dots, a_{Dj}\right); \quad j = 1,2, \dots, N$$

$$\boldsymbol{O_j} = \left(o_{1j}, o_{2j}, \dots, o_{Dj}\right); \quad j = 1,2, \dots, N$$

where N denotes the num of time or/and space grids, and D is the sum of d_m across all evaluated variables.

For model performance regarding multiple fields, the vector mean error (VME) measuring mean error is defined as:

$$VME = \sqrt{\sum_{i=1}^{D}(\bar{a}_i - \bar{o}_i)^2}, \ \bar{a}_i = \sum_{j=1}^{N} w_j \cdot a_{ij}, \ \bar{o}_i = \sum_{j=1}^{N} w_j \cdot o_{ij} \tag{S1}$$

where w_j represents the area weighting for grid *i* and its sum equals 1. Centered root-mean-square lengths (cRMSL) indicating anomaly amplitude for the model and the observation are, respectively, defined as:

$$cL_A = \sqrt{\sum_{i=1}^{D}\sum_{j=1}^{N} w_j \cdot (a_{ij} - \bar{a}_i)^2} \tag{S2}$$

and

$$cL_O = \sqrt{\sum_{i=1}^{D}\sum_{j=1}^{N} w_j \cdot (o_{ij} - \bar{o}_i)^2} \tag{S3}$$

The centered vector similarity coefficient (cVSC) measuring anomalous similarity between the model and the observation is defined as:

$$cR_v = \frac{\sum_{i=1}^{D}\sum_{j=1}^{N} w_j \cdot (a_{ij} - \bar{a}_i) \cdot (o_{ij} - \bar{o}_i)}{cL_A \cdot cL_o} \tag{S4}$$

Text S2

With the aid of Eq. (16), by replacing ME, SD, and CORR with VME, cRMSL, and cVSC, respectively, the modified version of KGE (KGE") for multiple fields is expressed as:

$$KGE'' = 1 - \sqrt{\frac{VME^2}{cRMSL_O^2} + (\frac{cRMSL_M}{cRMSL_O} - 1)^2 + (cVSC - 1)^2} \quad \text{(S5)}$$

Then, we can define the decomposition of (KGE"+1)^2 to three error components related to the mean error, anomaly amplitude error, and anomaly similarity error:

$$(KGE'' + 1)^2 = E_{Mean}^{KGE''} + E_{AnoA}^{KGE''} + E_{AnoS}^{KGE''} \quad \text{(S6)}$$

with:

$$E_{Mean}^{KGE''} = \frac{VME^2}{cRMSL_O^2}$$

$$E_{AnoA}^{KGE''} = (\frac{cRMSL_M}{cRMSL_O} - 1)^2$$

$$E_{AnoS}^{KGE''} = (cVSC - 1)^2$$

Text S3

Assume that there are two simulations of a scalar field derived from models A and B, and their overall performances by the normalized Decomposable Error Index (nDEI) can be expressed as:

$$nDEI_A^2 = E_{Mean}^{Nor,A} + E_{AnoA}^{Nor,A} + E_{AnoS}^{Nor,A} = \frac{ME_{A,O}^2}{SD_O^2} + (\frac{SD_A}{SD_O} - 1)^2 + 2 \cdot (1 - R_{A,O}) \quad \text{(S7)}$$

$$nDEI_B^2 = E_{Mean}^{Nor,B} + E_{AnoA}^{Nor,B} + E_{AnoS}^{Nor,B} = \frac{ME_{B,O}^2}{SD_O^2} + (\frac{SD_B}{SD_O} - 1)^2 + 2 \cdot (1 - R_{B,O}) \quad \text{(S8)}$$

where O represents the observation.

We defined the relative difference in the overall performance of model A against model B as:

$$RD_{nDEI} = \frac{nDEI_A - nDEI_B}{nDEI_A} \times 100\% \quad \text{(S9)}$$

Then, we conducted the attribution analysis for the inter-model difference in the overall performance from mean error, anomaly amplitude error, and anomaly similarity error. Here, we considered the inter-model nDEI^2 difference. The relative contributions of the three statistical errors to the DEI^2 difference are defined:

$$RC_{Mean}^{\Delta nDEI^2} = \frac{E_{Mean}^{Nor,A} - E_{Mean}^{Nor,B}}{nDEI_A^2 - nDEI_B^2} \times 100\% = \frac{\frac{ME_{A,O}^2}{SD_O^2} - \frac{ME_{B,O}^2}{SD_O^2}}{nDEI_A^2 - nDEI_B^2} \times 100\% \quad \text{(S10)}$$

$$RC_{AnoA}^{\Delta nDEI^2} = \frac{E_{AnoA}^{Nor,A} - E_{AnoA}^{Nor,B}}{nDEI_A^2 - nDEI_B^2} \times 100\% = \frac{(\frac{SD_A}{SD_O} - 1)^2 - (\frac{SD_B}{SD_O} - 1)^2}{nDEI_A^2 - nDEI_B^2} \times 100\% \quad \text{(S11)}$$

$$RC_{AnoS}^{\Delta nDEI^2} = \frac{E_{AnoS}^{Nor,A} - E_{AnoS}^{Nor,B}}{nDEI_A^2 - nDEI_B^2} \times 100\% = \frac{2 \cdot (1 - R_{A,O}) - 2 \cdot (1 - R_{B,O})}{nDEI_A^2 - nDEI_B^2} \times 100\% \quad \text{(S12)}$$

Among the three statistical errors, the error with the largest relative contribution to the inter-model DEI^2 difference emerges as the primary driver for the overall model performance difference. Beyond the evaluation of a scalar field, we can utilize Eqs. (S1–S6) to conduct a similar analysis for multiple fields, by replacing nDEI, ME, SD, and CORR with nDVEI, VME, cRMSL, and cVSC (Eqs. 10 and S1–S4), respectively.

Text S4

Following Text S1, for each of *M* variables that is scalar or vector, use Eqs. (A1–A5) and replace D with d_m to calculate the model performances in the corresponding aspects, which are the mean error (ME), the standard deviation (SD), and the correlation coefficient (R) (Zhang et al., 2021). Consequently, we can construct the relationship of model performances between multiple fields and individual variables as:

$$VME = \sqrt{\sum_{m=1}^{M} ME^{m\,2}} \tag{S13}$$

$$cL_A = \sqrt{\sum_{m=1}^{M} SD_A^{m\,2}} \tag{S14}$$

$$cL_O = \sqrt{\sum_{m=1}^{M} SD_O^{m\,2}} \tag{S15}$$

$$cR_v = \frac{\sum_{m=1}^{M}(R^m \cdot SD_A^m \cdot SD_O^m)}{\sqrt{\sum_{m=1}^{M} SD_A^{m\,2} \cdot \sum_i^M SD_O^{m\,2}}} \tag{S16}$$

where superscript *m* represents the *m*-th variable in the multivariable evaluation.

With the aid of Eqs. (2–5 and 10), we further decompose E_{Mean}, E_{AnoA}, and E_{AnoS} in terms of multiple fields as a function of those for individual variables:

$$E_{Mean} = \sum_m^M E_{Mean}^m \tag{S17}$$

$$\begin{aligned} E_{AnoA} &= cL_A^2 + cL_O^2 - 2 \cdot cL_A \cdot cL_o \\ &= \sum_m^M [SD_A^{m\,2} + SD_O^{m\,2} - 2 \cdot (SD_A^m \cdot SD_O^m) + 2 \cdot (SD_A^m \cdot SD_O^m)] - 2 \cdot cL_A \cdot cL_o \\ &= \sum_m^M E_{AnoA}^m + 2 \cdot \sum_m^M (SD_A^m \cdot SD_O^m) - 2 \cdot \sqrt{\sum_m^M SD_A^{m\,2} \cdot \sum_m^M SD_O^{m\,2}} \end{aligned} \tag{S18}$$

$$\begin{aligned} E_{AnoS} &= 2 \cdot cL_O^2 \cdot (1 - cR_v) \\ &= 2 \cdot \sum_{m=1}^{M} SD_O^{m\,2} - 2 \cdot \sqrt{\frac{\sum_{m=1}^{M} SD_O^{m\,2}}{\sum_i^M SD_A^{m\,2}}} \cdot \sum_{m=1}^{M} (R^m \cdot SD_A^m \cdot SD_O^m) \\ &= 2 \cdot \sum_{m=1}^{M} SD_O^{m\,2} - 2 \cdot \sum_{m=1}^{M} \left(SD_O^{m\,2} \cdot R^m\right) \\ &\quad + 2 \cdot \sum_{m=1}^{M} [SD_O^m \cdot R^m \cdot (SD_O^m - SD_A^m \cdot \sqrt{\frac{\sum_{m=1}^{M} SD_O^{m\,2}}{\sum_i^M SD_A^{m\,2}}})] \\ &= \sum_m^M E_{AnoS}^m + 2 \cdot \sum_{m=1}^{M} [SD_O^m \cdot R^m \cdot (SD_O^m - SD_A^m \cdot \sqrt{\frac{\sum_{m=1}^{M} SD_O^{m\,2}}{\sum_i^M SD_A^{m\,2}}})] \end{aligned} \tag{S19}$$

Table

Table S1 Model name, horizontal resolution, and institution of 26 CMIP6 models used in the study.

	Model Name	Horizontal Resolution	Institution
1	ACCESS-ESM1-5	1.88°×1.25°	Commonwealth Scientific and Industrial Research Organization (Australia)
2	AWI-CM-1-1-MR	0.94°×0.94°	Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (Germany)
3	BCC-CSM2-MR	1.13°×1.13°	Beijing Climate Center (China)
4	CAMS-CSM1-0	1.13°×1.13°	Chinese Academy of Meteorological Sciences (China)
5	CESM2-WACCM	1.25°×0.94°	National Center for Atmospheric Research (NCAR) (USA)
6	CESM2	1.25°×0.94°	
7	CIESM	1.25°×0.94°	Tsinghua University (China)
8	CanESM5	2.81°×2.81°	Canadian Centre for Climate Modelling and Analysis (Canada)
9	E3SM-1-1	1°×1°	Lawrence Livermore National Laboratory (USA)
10	EC-Earth3	0.7°×0.7°	European Center Earth Consortium (Europe)
11	FGOALS-f3-L	2.25°×2°	Chinese Academy of Sciences (China)
12	FGOALS-g3	1°×1°	
13	FIO-ESM-2-0	1.25°×0.94°	First Institute of Oceanography, State Oceanic Administration (China)
14	GFDL-ESM4	1.25°×1°	Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration (USA)
15	GISS-E2-1-G	2.5°×2°	Goddard Institute for Space Studies, National Aeronautics and Space Administration (USA)
16	GISS-E2-1-H	2.5°×2°	
17	INM-CM5-0	1.88°×1.25°	Institute for Numerical Mathematics, Russia Academy of Science (Russia)
18	IPSL-CM6A-LR	2.5°×1.16°	Institute Pierre Simon Laplace (France)
19	KACE-1-0-G	1.88°×1.25°	National Institute of Meteorological Sciences/ Korea Meteorological Administration (Korea)
20	MCM-UA-1-0	3.75°×2.25°	Department of Geosciences, University of Arizona (USA)
21	MIROC6	1.41°×1.41°	Japan Agency for Marine-Earth Science and Technology (Japan)
22	MPI-ESM1-2-HR	1.88°×1.88°	Max Planck Institute for Meteorology (Germany)
23	MPI-ESM1-2-LR	0.94°×0.94°	
24	MRI-ESM2-0	1.13°×1.13°	Meteorological Research Institute, Japan Meteorological Agency (Japan)
25	NESM3	1.88°×1.87°	Nanjing University of Information Science and Technology (China)
26	NorESM2-MM	1.25°×0.94°	Norwegian Climate Center (Norway)

Table S2 Information of evaluated variables in the study, including the unit and acronym. Vector wind consists of meridional and zonal winds.

	Variable name	**Unit**	**Acronym**
Surface variables	Near-surface temperature	degC	Tas
	Precipitation	mm/day	Pr
Air variables (850hPa, 500hPa, 200hPa)	Vector wind	m/s	UV
	Geopotential height	m	Z
	Air temperature	K	T
	Specific humidity	g/kg	Q

Table S3 Model name and horizontal resolution for 13 pairs of CMIP5 and CMIP6 models. In each pair, the first (second) line shows the CMIP5 (CMIP6) model.

	Name	**Horizontal Resolution**
ACCESS	CSIRO-Mk3-6-0	1.88°×1.88°
	ACCESS-CM2	1.88°×1.25°
BCC-CSM	BCC-CSM1-1	2.81°×2.81°
	BCC-CSM2-MR	1.13°×1.13°
CESM	CNRM-CM5	1.41°×1.41°
	CESM2	1.25°×0.94°
CanESM	CanESM2	2.81°×2.81°
	CanESM5	2.81°×2.81°
EC-Earth	EC-Earth2	1.125°×1.125°
	EC-Earth3	0.7°×0.7°
FGOALS-g	FGOALS-g2	3°×2.8°
	FGOALS-g3	2.25°×2°
GFDL-ESM	GFDL-ESM2M	2.50°×2.00°
	GFDL-ESM4	1.25°×1°
GISS-E2	GISS-E2-R	2.5°×2.0°
	GISS-E2-1-G	2.5°×2.0°
IPSL-CM	IPSL-CM5A-LR	2.5°×1.25°
	IPSL-CM6A-LR	2.5°×1.25°
INM	inmcm4	2°×1.5°
	INM-CM5-0	2°×1.5°
MIRCOC	MIRCO5	1.41°×1.41°
	MIRCO6	1.41°×1.41°
MPI-ESM	MPI-ESM-MR	1.875°×1.875°
	MPI-ESM1-2-HR	0.94°×0.94°
NorESM	NorESM1-M	2.3°×2.5°
	NorESM2-MM	1.25°×0.94°

Table S4 Information on reference datasets of precipitation, near-surface temperature, and air variables.

	Dataset	Horizontal and Temporal Resolution
Precipitation	Multi-Source Weighted-Ensemble Precipitation Version 2 (MSWEP V2)	0.1°×0.1° 3h
	Global Precipitation Climatology Centre (GPCC)	1°×1° monthly
	Climatic Research Unit Time-Series (CRU TS), v4.09	0.5°×0.5° monthly
Near-surface temperature	Berkeley Earth	1°×1° monthly
	CRU TS, v4.09	0.5°×0.5° monthly
	CPC GHCN_CAMS Gridded 2m Temperature (Land)	0.5°×0.5° monthly
Air variables	European Centre for Medium-Range Weather Forecasts Reanalysis 5 (ERA5)	0.25°×0.25° monthly

Table S5 The consistency between model rankings based on different summarized performance metrics, i.e., nDEI, nRMSE, and KGE", which is measured by the fraction of model pairs for which two metrics agree on the relative order. nDEI is the normalized decomposable error index, nRMSE is the normalized Root Mean Square Error, and KGE" is the modified Kling-Gupta Efficiency. Evaluated variables include precipitation (Pr), near-surface temperature (Tas), multi-level air temperatures (T), multi-level geopotential heights (Z), multi-level specific humidities (Q), and multi-level vector winds (UV).

(a) JJA

	Pr	**Tas**	**T**	**Z**	**Q**	**UV**
nDEI .vs. nRMSE	92%	95%	97%	99%	97%	96%
nDEI .vs. KGE"	90%	71%	86%	91%	80%	75%
KGE" .vs. nRMSE	93%	72%	85%	92%	81%	76%

(b) DJF

	Pr	**Tas**	**T**	**Z**	**Q**	**UV**
nDEI .vs. nRMSE	91%	99%	100%	100%	97%	95%
nDEI .vs. KGE"	85%	87%	87%	92%	87%	76%
KGE" .vs. nRMSE	89%	87%	87%	92%	88%	79%

Table S6 Model performance of 26 CMIP6 models in terms of mean error, standard deviation, and correlation coefficient in simulating climatological mean field of near-surface temperature in JJA throughout 1979–2005 over the CORDEX-EA region. The observed standard deviation of near-surface temperature is 7.47.

Model	Mean error	Standard deviation	Correlation coefficient
ACCESS-ESM1-5	1.10	7.28	0.95
AWI-CM-1-1-MR	1.10	7.26	0.94
BCC-CSM2-MR	0.09	7.84	0.95
CAMS-CSM1-0	0.51	7.35	0.97
CESM2-WACCM	1.42	7.23	0.97
CESM2	1.57	7.27	0.97
CIESM	2.32	7.12	0.95
CanESM5	0.28	8.48	0.92
E3SM-1-1	1.40	7.84	0.96
EC-Earth3	0.51	8.00	0.97
FGOALS-f3-L	0.18	7.94	0.97
FGOALS-g3	0.60	8.48	0.95
FIO-ESM-2-0	1.26	7.38	0.96
GFDL-ESM4	0.43	8.47	0.97
GISS-E2-1-G	0.76	8.45	0.94
GISS-E2-1-H	0.52	8.66	0.94
INM-CM5-0	0.13	6.42	0.94
IPSL-CM6A-LR	1.30	8.98	0.96
KACE-1-0-G	0.18	7.61	0.93
MCM-UA-1-0	2.90	6.64	0.83
MIROC6	4.17	8.98	0.85
MPI-ESM1-2-HR	0.93	7.54	0.96
MPI-ESM1-2-LR	0.32	7.10	0.93
MRI-ESM2-0	1.23	8.38	0.97
NESM3	0.69	7.07	0.91
NorESM2-MM	1.01	7.03	0.97

Table S7 Summary of statistical metrics and their calculation equations defined in this study.

(a) Statistical metrics for model evaluation of a scalar field

Long name	**Acronym**	**Equation**
Decomposable Error Index	DEI	$DEI = \sqrt{E_{Mean} + E_{AnoA} + E_{AnoS}}$
Penalty term for mean error in DEI	E_{Mean}	$E_{Mean} = ME^2$
Penalty term for anomaly amplitude error in DEI	E_{AnoA}	$E_{AnoA} = (SD_A - SD_O)^2$
Penalty term for anomaly similarity error in DEI	E_{AnoS}	$E_{AnoS} = 2 \cdot {SD_O}^2 \cdot (1 - R)$
Normalized DEI	nDEI	$nDEI = \sqrt{\frac{ME^2}{SD_O^2} + (\frac{SD_M}{SD_O} - 1)^2 + 2 \cdot (1 - R)}$
Relative contribution of mean error to DEI[2]	RC_{Mean}	$RC_{Mean} = \frac{ME^2}{ME^2 + (SD_A - SD_O)^2 + 2 \cdot {SD_O}^2 \cdot (1 - R)} \times 100\%$
Relative contribution of anomaly amplitude error to DEI[2]	RC_{AnoA}	$RC_{AnoA} = \frac{(SD_A - SD_O)^2}{ME^2 + (SD_A - SD_O)^2 + 2 \cdot {SD_O}^2 \cdot (1 - R)} \times 100\%$
Relative contribution of anomaly similarity error to DEI[2]	RC_{AnoS}	$RC_{AnoS} = \frac{2 \cdot {SD_O}^2 \cdot (1 - R)}{ME^2 + (SD_A - SD_O)^2 + 2 \cdot {SD_O}^2 \cdot (1 - R)} \times 100\%$

(b) Statistical metrics for model evaluation of vector/multiple fields

Long name	**Acronym**	**Equation**
Decomposable Vector Error Index	DVEI	$DVEI = \sqrt{E_{Mean} + E_{AnoA} + E_{AnoS}}$
Penalty term for mean error in DVEI	E_{Mean}	$E_{Mean} = VME^2$
Penalty term for anomaly amplitude error in DVEI	E_{AnoA}	$E_{AnoA} = (cL_A - cL_O)^2$
Penalty term for anomaly similarity error in DVEI	E_{AnoS}	$E_{AnoS} = 2 \cdot {cL_O}^2 \cdot (1 - cR_v)$
Normalized DVEI	nDVEI	$nDVEI = \sqrt{\frac{VME^2}{cL_O^2} + (\frac{cL_M}{cL_O} - 1)^2 + 2 \cdot (1 - cR_v)}$
Relative contribution of mean error to DVEI[2]	RC_{Mean}	$RC_{Mean} = \frac{VME^2}{VME^2 + (cL_A - cL_O)^2 + 2 \cdot {cL_O}^2 \cdot (1 - cR_v)} \times 100\%$
Relative contribution of anomaly amplitude error to DVEI[2]	RC_{AnoA}	$RC_{AnoA} = \frac{(cL_A - cL_O)^2}{VME^2 + (cL_A - cL_O)^2 + 2 \cdot {cL_O}^2 \cdot (1 - cR_v)} \times 100\%$
Relative contribution of anomaly similarity error to DVEI[2]	RC_{AnoS}	$RC_{AnoS} = \frac{2 \cdot {cL_O}^2 \cdot (1 - cR_v)}{VME^2 + (cL_A - cL_O)^2 + 2 \cdot {cL_O}^2 \cdot (1 - cR_v)} \times 100\%$

Figure

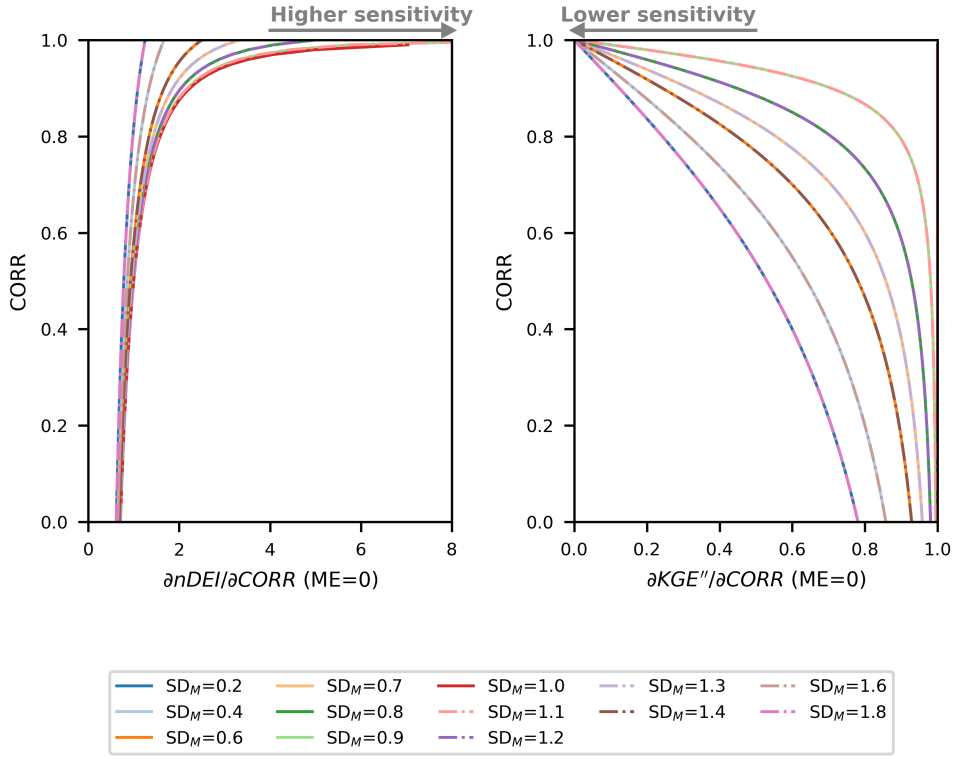


Figure S1: Sensitivity of nDEI and KGE" to the CORR, assuming the mean error is zero.

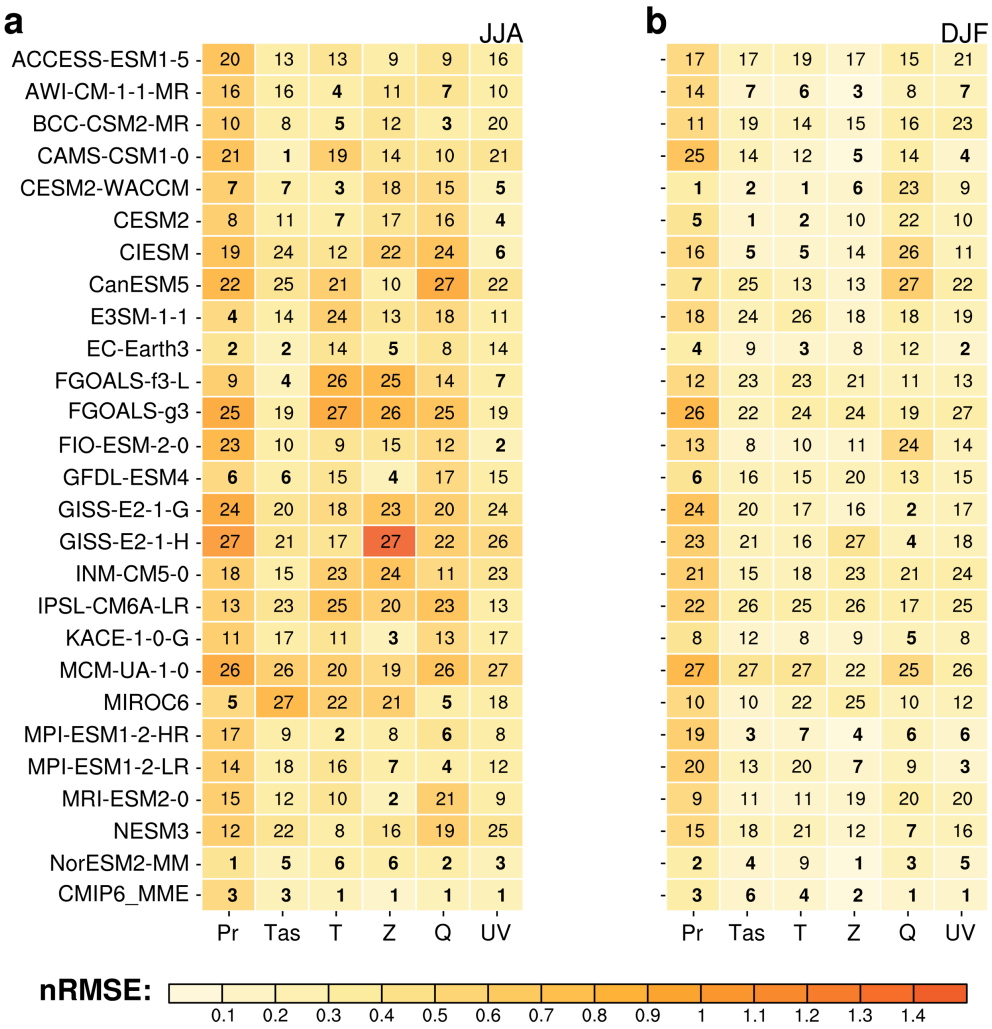


Figure S2: Same as Figs. 3a and b, but for the normalized Root Mean Square Error (nRMSE). As T, Z, Q, and UV are multiple fields, the calculation of nRMSE for these variables uses Eq. 15, but replaces ME, SD, and CORR with corresponding metrics for multiple fields, which are vector mean error (VME), centered root mean square length (cRMSL), and centered vector similarity correlation (cVSC), respectively.

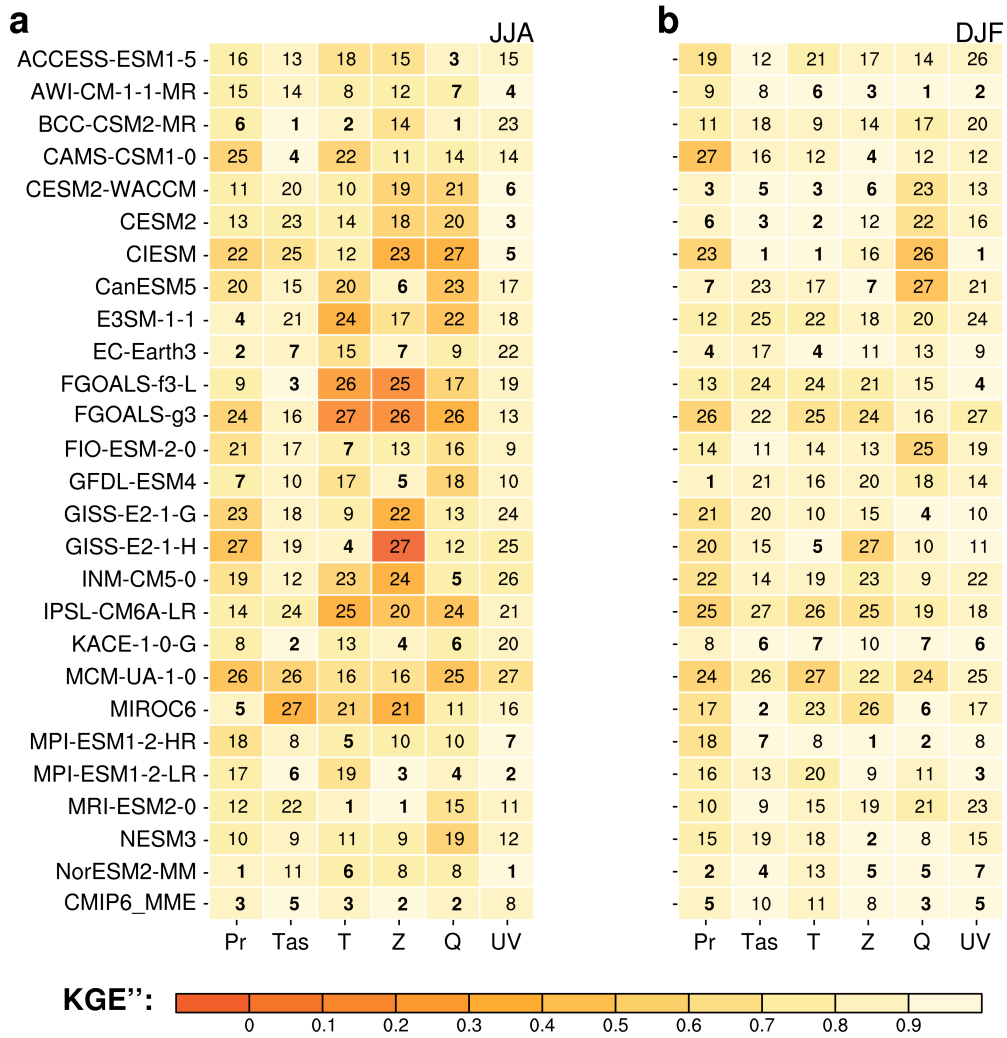


Figure S3: Same as Figs. 3a and b, but for the modified Kling-Gupta Efficiency (KGE"). As T, Z, Q, and UV are multiple fields, the calculation of nRMSE for these variables uses Eq. 16, but replaces ME, SD, and CORR with corresponding metrics for multiple fields, which are vector mean error (VME), centered root mean square length (cRMSL), and centered vector similarity correlation (cVSC), respectively.